Ano 26 Nº 38
Dez de 2008

ESTUDOS

AVALIAÇÃO DO ENSINO SUPERIOR BRASILEIRO: MODELOS, DIVULGAÇÃO DOS RESULTADOS E NORMAS REGULAMENTADORAS

E82 Estudos : Revista da Associação Brasileira de Mantenedoras de Ensino Superior : Avaliação do ensino superior Brasileiro : modelo, divulgação dos resultados e normas regulamentadoras / Associação Brasileira de Mantenedora de Ensino superior. – Ano 26, n. 38 (Nov. 2008). – Brasília : Associação Brasileira de Mantenedoras de Ensino Superior, 2008-
v. ; 30 cm.

Descrição baseada em: Ano 26, n. 38 (nov. 2008)

1. Ensino superior – avaliação. 2. Ensino superior – normas. 3. Ensino superior – legislação. I. Associação Brasileira de Mantenedoras de Ensino Superior. II. Título: Revista da Associação Brasileira de Mantenedoras de Ensino Superior. III. Título: Avaliação do ensino superior brasileiro. IV. Título: Modelos, divulgação dos resultados e normas regulamentadoras.

CDU 378(05)

Associação Brasileira de Mantenedoras de Ensino Superior
SCS Quadra 07 – Bloco “A”
Torre Pátio Brasil Shopping - Sala 526
70 330-911 - Brasília - DF
Tel.: (61) 3322-3252 Fax: (61) 3224-4933
E-mail: abmes@abmes.org.br
Home page: http://www.abmes.org.br

Ano 26 Nº 38
Dez de 2008

ESTUDOS

SUMÁRIO

APRESENTAÇÃO

GABRIEL MARIO RODRIGUES*

> "O problema com a avaliação do ensino superior brasileiro não é que ela exista, mas a forma como é feita e como os resultados são divulgados." (Simon Schwartzman, 2008)

A comunidade acadêmica de todo o País – e não apenas o segmento privado – teceu duras críticas ao sistema de avaliação das instituições de ensino superior (IES) e de cursos implementado pelo Ministério da Educação (MEC). Um dos principais pontos de divergência é a divulgação do *ranking* de instituições, feito a partir dos índices divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC).

Dirigentes das IES brasileiras públicas e privadas consideram a avaliação externa muito importante mas enfatizam que a formulação e a implementação de indicadores – como é o caso do Conceito Preliminar de Curso (CPC) precisam ser revistas. Na fase da avaliação, em que a instituição têm uma nota que ainda não é definitiva e que depende de uma avaliação *in loco,* a divulgação dos resultados pode gerar equívocos e graves prejuízos para as instituições.

Para discutir essas questões a ABMES realizou, no dia 9 de setembro de 2008, em Brasília, o seminário "Questões polêmicas da avaliação externa – o Enade, o IDD e o CPC". Participaram

* Presidente da Associação Brasileira de Mantenedoras de Ensino Superior (ABMES) e reitor da Universidade Anhembi Morumbi

como conferencistas Reynaldo Fernandes, presidente do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC); Simon Schwartzman, presidente do Instituto de Estudos do Trabalho e Sociedade (Iets); Celso da Costa Frauches, consultor sênior do Instituto Latino Americano de Planejamento Educacional (Ilape); e, Raulino Tramontin, consultor universitário.

Reynaldo Fernandes, ao explicar o desenho geral do sistema de avaliação, usou como pano de fundo para a sua explanação o papel do Inep "como órgão que tem por objetivo aferir a qualidade das instituições" e evitou entrar no campo da regulação. Para ele a avaliação padece de um pecado original – "estar associada à regulação."

Para rebater as críticas sobre a criação pelo MEC de índices e conceitos por meio de Portarias, Reynaldo afirmou que a Lei n.º 10.861/2004, que instituiu o Sistema Nacional de Avaliação da Educação Superior (Sinaes), permite diferentes formas de operacionalização, cujos resultados se destinam a diferentes públicos externo e interno às IES.

O presidente do Inep descreveu tecnicamente os três indicadores da avaliação – o Exame Nacional de Desempenho do Estudante (Enade) – "indicador de resultados"; o Índice de Diferença entre os Desempenhos Observado e Esperado (IDD), – "contribuição de um curso superior"; e, o Conceito Preliminar de Curso (CPC). Explicou que o CPC foi criado com o objetivo de reduzir as visitas *in loco* e de ser um "guia de orientação" para as comissões de especialistas, tendo em vista a impossibilidade concreta do Inep de avaliar todos os cursos de graduação do País. Nesse sentido, de acordo com Fernandes, os cursos avaliados com notas 3 a 5 não precisam ser visitados pelas Comissões.

Schwartzman considera que "embora preliminares, e aparentemente sujeitos à revisão, esses conceitos foram amplamente divulgados pela imprensa, afetando a reputação e provocando a reação indignada de muitas pessoas e instituições. Para ele – que é favorável à avaliação e que considera os dados estatísticos essenciais ao processo – "o problema com a avaliação do ensino superior brasileiro não é que ela exista, mas a forma como é feita e como os resultados são divulgados." O MEC busca avaliar um sistema extremamente heterogêneo com a implantação de um modelo único para todas as instituições – "uma camisa de força" – em prejuízo da diversidade que deveria se respeitada e valorizada.

O CPC de acordo com Schwartzman, é uma construção estatística, baseada em uma série de aproximações e pressupostos não explicitados. "Tal conceito pode fazer sentido como exercício de análise e até mesmo para ajudar nas decisões internas do MEC mas nunca poderia ser difundido publicamente como avaliação de qualidade de cursos, mesmo com o título de preliminar".

Schwartzman apresentou algumas idéias para aperfeiçoar o sistema de avaliação e, com isso, torná-lo representativo e legítimo: a) a criação de uma agência de regulação do ensino superior autônoma;

b) a descentralização das avaliações; c) a criação de um mercado competitivo de qualidade; d) a substituição das notas ou conceitos por certificações; e, e) o desenvolvimento de padrões claros de competência para as diferentes áreas do conhecimento.

Raulino Tramontin destacou que o Estado, ao exercer as funções de órgão regulador e avaliador, gera problemas sérios para as IES. O processo regulatório, que tem como base os dados do Inep, é alvo de discordâncias e de polêmicas. Para Tramontin, as interpretações da Lei do Sinaes e da Lei n.º 9.394/1996 (Lei de Diretrizes e Bases da Educação, LDB), feitas por quem está no poder, formam um dueto que possibilita uma série de combinações que poucos entendem a não ser o que a imprensa – privilegiada por receber em primeiro lugar as informações – interpreta e, com isso, contribui para a formação da opinião pública".

Tramontin afirmou que as IES particulares, que sempre se posicionaram favoravelmente à avaliação, têm o direito de questionar o conteúdo e a legalidade das normas regulamentadoras. Ele defendeu a abertura do diálogo entre o MEC e as IES; a reestruturação das comissões de especialistas com base na ética e no tratamento justo com as instituições; a avaliação baseada no Plano de Desenvolvimento Institucional (PDI): e, o cuidado em atender aos dispositivos da Lei do Sinaes.

Celso Frauches analisou o sistema de avaliação do ponto de vista legal sem se deter nas "fórmulas mágicas" porque "só o MEC pode explicá-las." Celso considera que o IDD e o CPC são ilegais pelo fato de terem sido criados por meio de Portarias. Para ele, a Portaria n. º 40/2007 (renovação de reconhecimento de cursos de graduação) "regulamenta tudo e é um verdadeiro código que fere o artigo 209 da Constituição Federal (CF) e o art.46 da LDB". Disse ainda que os instrumentos de avaliação dos cursos de Direito e de Medicina foram feitos para impedir a criação de novos cursos, fato que demonstra o grande poder das corporações no interior do MEC.

Em nenhum momento Celso questionou a competência legal do MEC de avaliar as IES. Para ele "as normas, procedimentos e instrumentos de avaliação devem ter origem em lei, aprovada no Congresso Nacional, após amplo debate com a sociedade. Não se pode aceitar como legítimo um processo avaliativo sem qualquer respaldo legal e apoiado em fórmulas mágicas, que agridem os mais elementares critérios de avaliação de qualidade da educação superior".

Após a realização do seminário, a ABMES compôs um grupo de trabalho (GT) para discutir as questões debatidas. O documento resultante do GT – "O Conceito Preliminar de Cursos e o Índice Geral de Cursos da Instituição de Educação Superior no Contexto do Sistema Nacional de Avaliação da Educação Superior" – tem como objetivo "demonstrar a inconsistência e a ilegalidade das atuais regras definidas pelo MEC. As instituições particulares sempre estiveram dispostas a cumprir o que dispõe o inciso II do artigo 209 da Constituição Federal que define como condição para o funcionamento do ensino superior a autorização e a avalição de qualidade pelo Poder Público (...). Além do cumprimento do aspecto legal, as IES acreditam na importância da avaliação evidenciada

pelas inúmeras mobilizações, estudos e seminários realizados por suas representações com o objetivo de assegurar o cumprimento do papel institucional às disposições do processo de avaliação do MEC. Os argumentos apresentados no documento do GT evidenciam que as atuais medidas tomadas pelo MEC apresentam fragilidades técnicas e descumprem as disposições do Sinaes e da Lei de Diretrizes e Bases."

O documento do GT reforça a crença de que "os processos avaliativos, como dispostos na Lei do Sinaes, são a garantia de qualidade do sistema de educação superior, pois superam a lógica da avaliação fragmentada, classificatória e de verificação e promove uma análise sistemática e integrada dos processos avaliativos de instituições e de cursos e do desempenho dos estudantes. Interpretar adequadamente a Lei e concretizar seus princípios é atender à diretrizes traçadas no Plano Nacional de Educação que enfatiza a reformulação do rígido sistema de controles burocráticos."

Para tratar dessas e de outras questões, a ABMES decidiu organizar a presente edição da revista Estudos com a íntegra dos textos de Simon Schwatzman, de Raulino Tramontin, de Celso da Costa Frauches e do GT – referências importantes para refletir e debater os processos de avaliação interna e externa das instituições de ensino supeerior.

Brasília, 2 de dezembro de 2008.

O "CONCEITO PRELIMINAR" E AS BOAS PRÁTICAS DE AVALIAÇÃO DO ENSINO SUPERIOR

SIMON SCHWARTZMAN*

O Ministério da Educação divulgou, no dia 6 de agosto de 2008, o até então desconhecido "conceito preliminar de curso" que classificou 508 dos 2.028 cursos superiores avaliados pelo Exame Nacional de Avaliação do Desempenho de Alunos (Enade) de 2007 como de qualidade insuficiente: 444 da rede privada (19,5% do setor), e 64 da rede pública (12,2%). Embora preliminares e aparentemente sujeitos a revisão, estes conceitos foram amplamente divulgados pela imprensa, afetando a reputação e provocando a reação indignada de muitas pessoas e instituições.

Existem de fato muitos cursos superiores de má qualidade no Brasil, públicos e privados, que precisam ser avaliados de forma externa e independente. A avaliação, quando bem feita, informa o público sobre cursos que devem ser buscados ou evitados, e estimula as instituições a melhorar seu desempenho. A auto-avaliação não é suficiente, porque ela não produz resultados comparáveis, e são geralmente defensivas. O problema com a avaliação do ensino superior brasileiro não é que ela exista, mas a forma como ela é feita, e como os resultados são divulgados. O objetivo deste texto é examinar com algum detalhe a forma como este "conceito preliminar" foi desenvolvido e utilizado, à luz das boas práticas internacionais de avaliação externa da educação superior.

O "conceito preliminar de avaliação" dos cursos de nível superior foi elaborado e difundido pelo Ministério da Educação com o propósito de utilizá-lo para decidir se os cursos necessitam ou não de

* Membro do Conselho de Administração e pesquisador do Instituto de Estudos do Trabalho e Sociedade (Iets). simon@iets.inf.br

avaliação *in loco*, conforme o previsto na legislação que reorganizou o sistema de avaliação da educação superior no país,[1] para serem ou não autorizados a funcionar. Segundo a Portaria Normativa n.º 4, de 05 de agosto de 2008,

> Art. 2.º Os cursos que tenham obtido conceito preliminar satisfatório ficam dispensados de avaliação *in loco* nos processos de renovação de reconhecimento respectivos.
>
> §1.º Considera-se conceito preliminar satisfatório igual ou superior a três.
>
> §2.º Os processos de renovação de reconhecimento dos cursos que tenham obtido conceito 5 (cinco), em tramitação nos sistemas Sapiens ou e-MEC, serão encaminhados à Secretaria competente, para expedição da Portaria de renovação de reconhecimento.

A avaliação da educação superior no Brasil e as boas práticas de avaliação externa

A idéia de que os cursos superiores, assim como de outros níveis, devem ser submetidos a avaliações externas é hoje bastante consensual em todo o mundo, embora a maneira pela qual esta avaliação deva ser feita e a forma pela qual estes resultados devem ser divulgados, sejam ainda sujeitos a controvérsia.

Em um mundo ideal, as instituições educativas deveriam ter total autonomia para decidir quando e o que ensinar, e os alunos deveriam ter liberdade para escolher os cursos que considerassem melhores e mais adequados para si. No mundo real, os estudantes não têm informações adequadas a respeito dos cursos disponíveis, as instituições de ensino fazem uso de recursos e mandatos públicos (financiamentos diretos para as instituições públicas, isenções fiscais para as privadas, e autorização para conceder diplomas de validade legal para ambas) sobre os quais devem prestar contas, e é natural que cada instituição tenda a valorizar sua própria atuação.

A avaliação externa, quando bem conduzida, produz uma série de resultados importantes. Permite que as instituições de ensino se aperfeiçoem, a partir da comparação entre seu desempenho e o de outras instituições similares. Informa ao público – estudantes e suas famílias, futuros empregadores – a respeito da qualidade dos diversos cursos; permite que o setor público possa direcionar seus recursos da melhor maneira possível; e permite também que os cursos em condições de desempenho inaceitáveis sejam identificados e fechados ou levados a se ajustar.

Diferentes países adotam sistemas distintos de avaliação, e não existe um formato que possa ser considerado claramente superior aos demais.[2] Existem, no entanto, algumas características que distinguem as boas e más práticas destas avaliações, algumas relativas aos objetivos das avaliações,

---

1 Lei n.º 10.861, de 14 de abril de 2004, que institui o Sistema Nacional de Avaliação da Educação Superior (Sinaes).

2 Para um amplo panorama dos sistemas de avaliação da educação superior no mundo, ver Billing, D. (2004). International Comparisons and Trends in External Quality Assurance of Higher Education: Commonality or Diversity? *Higher Education*, 47(1), 113-137.

outras relativas à maneira pela qual os sistemas de avaliação são institucionalizados e utilizados, e outras relativas aos aspectos mais técnicos do uso de testes e outras estatísticas.

O que avaliar?

Dizer que o objetivo da avaliação é identificar as instituições e cursos de boa e má qualidade não é suficiente. Todas as pessoas – professores, alunos, pais, diretores, governantes – têm idéias próprias sobre o que é a qualidade, mas que nem sempre são compartidas pelos demais. Quais são as prioridades? Que tipo de competências e informações os alunos deveriam obter nos cursos? Estas expectativas de competência e informação devem ser iguais para todos os estudantes, ou devem variar conforme suas condições pessoais, motivações e interesses? Quais são os mínimos aceitáveis, e o que deve ser considerado inadmissível nas diversas carreiras?

No passado, quando a educação superior era limitada a uma pequena elite, estas questões eram definidas de forma implícita pelas pessoas de maior prestígio e reconhecimento das diversas áreas de conhecimento. Hoje, com sistemas altamente diferenciados e milhões de estudantes, estes critérios implícitos precisam vir à luz. No Brasil, ainda prevalece a idéia de que todo o ensino superior deve ser dado em universidades públicas, que os professores devem ser pesquisadores e trabalhar em regime de tempo integral, e que os estudantes devem também se dedicar primordialmente ao estudo e ser avaliados pelos critérios dos professores as instituições mais privilegiadas. E, no entanto, a maioria das instituições pertence ao segmento privado, a maioria dos professores só pratica o ensino, a maioria dos alunos estuda à noite e trabalha durante o dia, e poucos chegam ao ensino superior com condições de atender aos requisitos acadêmicos dos cursos mais exigentes.

Em algumas áreas, como as da saúde, engenharia, direito e educação, em que os formados devem ser capazes de desempenhar atividades que colocam em risco a vida, o patrimônio e a formação das pessoas, deve haver critérios mínimos de qualificação, abaixo dos quais a certificação para o exercício profissional não deveria ser concedida. Mesmo nestas áreas, no entanto, existem muitas variações de especialidade, e a tendência é a de criação de sistemas de certificação profissional individualizada para as diferentes especialidades, cujo nível de exigência pode variar. A maior parte dos estudantes de nível superior não está nestes cursos, mas em outros como administração, economia e comunicação, onde busca uma formação adicional e maior acesso ao mercado de trabalho, sem o requisito de uma formação especializada, que cada vez mais se dá na pós-graduação.

Esta situação é claramente incompatível com os sistemas de avaliação unidimensionais, que aplicam uma mesma métrica para avaliar todos os cursos ou instituições pelo mesmo critério. O Exame Nacional de Cursos (ENC, o "Provão") instituído pelo Ministro Paulo Renato de Souza, teve a vantagem de colocar a questão da qualidade da educação superior na pauta de discussão,[3] mas padeceu deste problema, pois avaliava a qualidade dos cursos pelo desempenho dos alunos que concluíam os cursos em um exame nacional, independentemente das diferentes orientações ou

preferências das instituições ou das características que os alunos traziam para o ensino superior em função de sua educação prévia. O Exame Nacional de Desempenho de Estudantes (Enade), em sua concepção inicial, buscou corrigir um destes problemas, ao avaliar os alunos iniciantes e concluintes, mas, como veremos mais abaixo, esta intenção acabou se frustrando pela maneira em que o exame foi implementado; e não avançou nada na direção de desenvolver critérios múltiplos de avaliação para instituições com objetivos diferentes.

Criar sistemas múltiplos e flexíveis de avaliação é difícil, porque os avaliadores tendem a ter critérios próprios e unidimensionais. Uma possibilidade é abrir espaço para a existência de múltiplas agências avaliadoras, devidamente credenciadas, às quais as instituições de ensino superior se filiem. México e Chile estabeleceram sistemas deste tipo, e as avaliações nos Estados Unidos são também descentralizadas e plurais. Outra alternativa é não pretender ordenar os cursos conforme uma métrica única, mas, simplesmente, credenciá-los ou não, em função do atendimento de certos critérios mínimos de desempenho, deixando a questão da certificação profissional nas mãos de agências e sistemas de certificação das diferentes profissões.

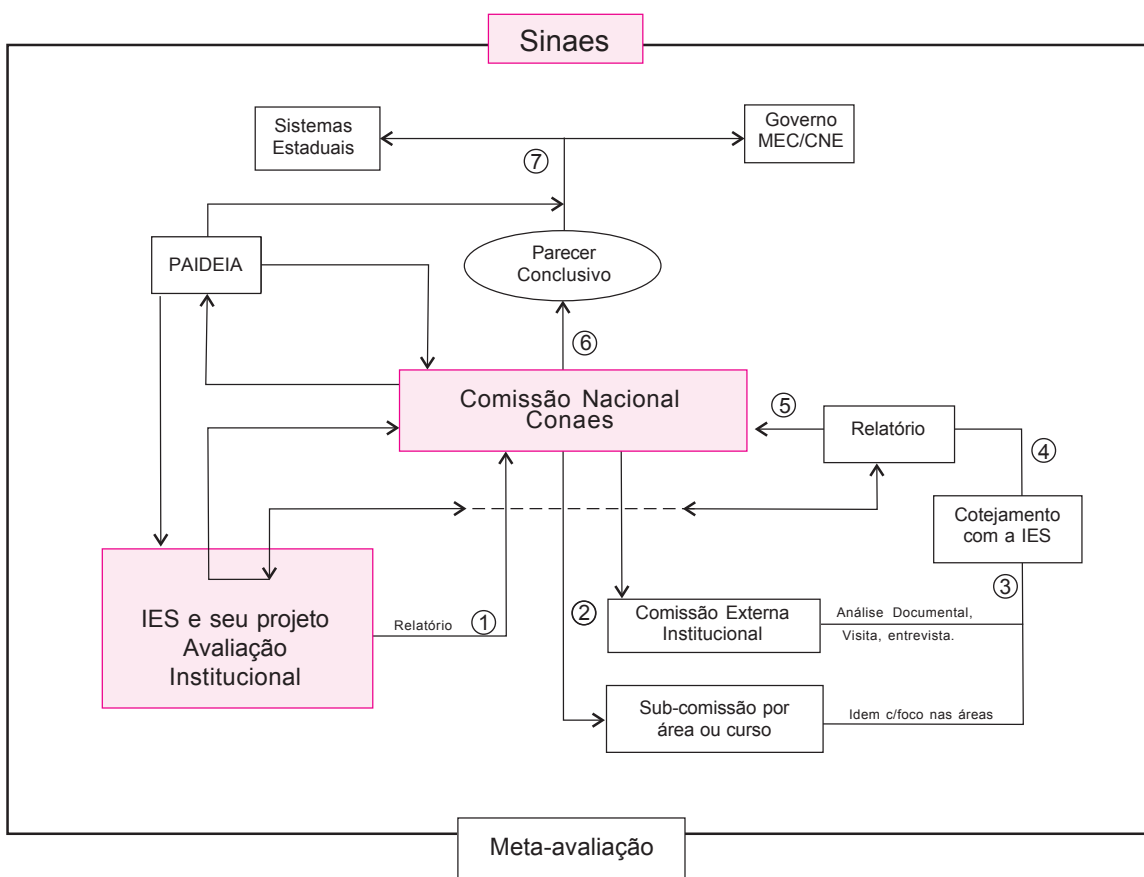
Quem avalia?

Avaliações são sempre o produto de um julgamento feito por pessoas. Não existem avaliasções "objetivas", porque provas, testes e indicadores são sempre definidos por pessoas que decidem o que deve ser medido e testado. Isto não significa que testes e medidas estatísticas não tenham importância. Dados como o despenho dos alunos em exames comparáveis, a posição dos ex-alunos no mercado de trabalho e em cursos avançados de pós-graduação, a qualificação profissional dos professores, a opinião dos alunos e professores sobre suas instituições e cursos, são indicadores importantes que podem ajudar os avaliadores a formar e balizar seus juízos. Existe uma grande diferença, no entanto, entre o uso de indicadores como elementos de diagnóstico e a substituição do juízo dos avaliadores por um índice ou índices calculados mecanicamente. Isto é lugar comum na medicina, aonde os médicos fazem uso de testes e avaliações estatísticas de seus pacientes, mas sempre devem tomar uma decisão individual, sob sua responsabilidade, a respeito de sua interpretação e das ações recomendadas.

Avaliações externas devem ser feitas por avaliadores, também externos e independentes, que não tenham envolvimento nem conflitos de interesses com as instituições avaliadas. Além da independência, as avaliações externas trazem para dentro dos cursos e instituições avaliadas a oportunidade de entrar em contato com pessoas experientes em suas áreas e são levadas a se comparar com outras, saindo, portanto, do isolamento e da auto-suficiência que muitas vezes predominam. Auto-avaliações são indispensáveis em qualquer instituição, e, além disto, permitem processar, reagir e se

[3] A necessidade da avaliação já estava presente nos trabalhos da Comissão Nacional de Avaliação Superior e das propostas do Grupo Executivo da Reforma da Educação Superior (Geres), arquivadas pelo Ministro Marco Maciel no governo José Sarney. Os documentos da Comissão e do Geres estão disponíveis em: http://www.schwartzman.org.br/simon/comissao.htm http://www.schwartzman.org.br/simon/pdf/geres.pdf

for o caso contestar as avaliações externas, mas não produzem resultados que sejam transparentes, visíveis e comparáveis para a sociedade mais ampla.

Em vários países, as avaliações são executadas por agências governamentais. É mais recomendável, no entanto, que elas sejam promovidas por instituições independentes, que não possam ser afetadas por eventuais câmbios de ministros ou governantes. O Brasil tem uma história de criação de colegiados aparentemente independentes, como o antigo Conselho Federal (e depois Nacional) de Educação, mas que nunca tiveram instrumentos adequados para implementar políticas próprias de avaliação. O Conselho Nacional de Avaliação do Ensino Superior (Conaes) deveria ser também um órgão colegiado independente, mas na realidade tem, entre seus membros, cinco representantes do governo, três representantes de interesses organizados do ensino superior (docentes, pessoal administrativo e estudantes) e somente cinco membros independentes, "representantes com notório saber científico, filosófico e artístico", de livre nomeação do Ministro da Educação. O Conaes foi responsável pela criação de um ambicioso e complexo sistema de avaliação do ensino superior brasileiro – o Sinaes – abrangente e participativo em suas intenções.[4]

[4] A descrição detalhada das intenções e implementação do Sinaes e uma visão das experiências anteriores de avaliação do ensino superior brasileiro podem ser vistas em Conaes. Sinaes – Sistema Nacional de Avaliação da Educação Superior: da concepção à regulamentação (2a ed.). Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2004).

A principal crítica que se pode fazer ao Sinaes, em termos gerais, é que o sistema proposto de avaliação institucional e de cursos é demasiado complexo (como se pode ver no gráfico acima, para as avaliações institucionais) e muito além dos recursos humanos e materiais disponíveis no Ministério da Educação. Na prática, a única coisa que funcionou, e que a sociedade tomou conhecimento, foi o Enade, implementado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) de forma independente. Neste processo, perdeu-se um elemento central do processo de avaliação, que deve culminar necessariamente em um parecer conclusivo por profissionais de grande legitimidade em sua área atuação. Isto claramente não ocorre quando o Ministério da Educação distribui "conceitos provisórios" produzidos administrativamente.

Como avaliar?

Qualquer professor sabe como preparar uma prova, mas a avaliação educacional sistemática é hoje uma área altamente especializada e sujeita a permanentes debates e revisões. Do ponto de vista técnico, as avaliações precisam explicitar com clareza o que está sendo medido, desenvolver indicadores para fazer a mensuração, garantir que de fato estes indicadores medem o que deveriam medir, assegurar que as escalas cubram a grande variedade das pessoas sendo avaliadas, e que os resultados de avaliações feitas em grupos diferentes e em diferentes pontos no tempo sejam comparáveis. Sistemas de avaliação tecnicamente bem-estruturados, como o Sistema de Avaliação de Educação Básica (Saeb), fazem uso da chamada "teoria dos itens de resposta" (item *response theory*) segundo a qual diferentes itens das provas possuem relações matematicamente definidas com dimensões latentes que se pretende avaliar. Isto permite medir as mesmas competências por baterias distintas de itens, com níveis diferentes de dificuldade, mas dentro de uma mesma escala de medição.

Além dos aspectos técnicos, é fundamental definir com clareza o que se pretende medir. A principal alternativa é entre a mensuração de conhecimentos e informações e a mensuração de competências. A maneira mais tradicional de avaliar é verificar se o avaliado é capaz de reproduzir aquilo que o professor lhe ensinou, o que, no extremo, é uma medida de sua capacidade de memorização. A tendência mais moderna é avaliar as competências – se ele é capaz de utilizar os conhecimentos e a educação que recebeu para lidar com situações e problemas novos, que exijam iniciativa e criatividade.[5] Na prática, as competências dependem também de conhecimentos, e por isto a linha divisória entre estas duas modalidades não é clara. As avaliações por competência têm sido criticadas muitas vezes por tentar fragmentar o conhecimento em dimensões isoladas, enquanto que as avaliações tradicionais são criticadas pela falta de clareza em relação ao que está sendo avaliado, e por reproduzir, simplesmente, as práticas pedagógicas dos professores.

---

5 O *Tuning Project,* criado para examinar a compatibilidade da formação proporcionada pelos diversos países europeus que participam do Processo de Bologna, define competências, consideradas o objetivo central dos programas educativos, como a dynamic combination of knowledge, understanding, skills and abilities. Veja a respeito http://tuning.unideusto.org/tuningeu/, na parte de competências.

O desenvolvimento de testes de avaliação se torna ainda mais difícil porque requer uma interação forte e permanente entre os especialistas em estatística e os especialistas das diversas competências que se pretende avaliar. São os especialistas de conteúdo que desenvolvem itens e que determinam a pertinência das dimensões avaliadas. Existem procedimentos estabelecidos por meio dos quais as avaliações são desenvolvidas com a participação de especialistas em avaliação e especialistas e representantes das áreas avaliadas, de forma tal que os testes sejam validados tanto do ponto de vista estatístico quanto de seu conteúdo, e seus resultados compreendidos e aceitos como válidos pela comunidade especializada.

Além destas questões, os especialistas em educação costumam distinguir as avaliações e medidas feitas com objetivos estatísticos, para entender situações e tendências de natureza geral, e as avaliações individuais, que afetam os interesses e as oportunidades de pessoas e instituições específicas. Em inglês, elas são denominadas *low stake* e *high stake asssessments*, que costumam ser denominados no Brasil como "avaliação com dentes", ou avaliação com conseqüências. O Saeb, e seus similares, são avaliações *low stakes*, porque seus resultados não afetam os estudantes que participam da avaliação, diferentemente do Exame Nacional do Ensino Médio (Enem), que é um exame de *high stakes*, porque influencia a chance do estudante de entrar para o ensino superior.

O Enade ocupa uma posição estranha em relação a isto. Para os estudantes, é de baixa conseqüência, porque seu desempenho não é registrado em seu currículo. Para as instituições, no entanto, as conseqüências são importantes, porque a média dos resultados será divulgada e afetará o prestígio e eventualmente o credenciamento dos cursos. Se os estudantes estiverem motivados em assegurar o prestígio e o reconhecimento de seus cursos, eles se esforçarão mais para se sair bem nas provas; se não, podem entregar a prova em branco. Além disto, as provas com conseqüências tendem a ser mais complexas e abrangentes, enquanto que as provas de baixas conseqüências, feitas para fins meramente estatísticos, tendem a ser mais ligeiras. Ao tentar fazer as duas coisas ao mesmo tempo, o Enade pode não estar fazendo bem nenhuma das duas.

Abrindo a caixa-preta do Conceito Preliminar

Estas considerações iniciais permitem olhar mais em detalhe como o "conceito preliminar" foi elaborado, e averiguar se ele está de acordo com as boas práticas existentes. Embora o Ministério da Educação não tenha divulgado todos os passos que o levaram ao resultado final, é possível tentar reconstruí-los, a partir de vários documentos do Inep, alguns publicados recentemente, outros anteriores, que já não estão disponíveis no *site* do Ministério.

Em uma nota técnica, o Ministério da Educação indica que o conceito preliminar resulta da combinação de três componentes, o resultado do Enade, o Indicador de Diferença de Desempenho (IDD) e

os “insumos”, com os pesos, respectivamente, de 40%, 30% e 30% no conceito final.[6] Cada um destes componentes é transformado em uma escala de 5 pontos, que é a também a escala final do CPC. A fórmula é a seguinte:

$$CPC_i = (\rho 1 * nota_Enade) + (\rho 2 * nota_IDD_i) + (\rho 3 * nota_Insumos_i)$$

Em linguagem corrente, o conceito preliminar é o resultado da soma dos resultados do Enade, do IDD e dos Insumos, cada qual com um peso específico – p1, p2, p3 - que somam 100%. Antes de entrar no detalhe da construção destes índices, é necessário perguntar de onde vieram estes pesos. Porque a do Enade só vale 40%, e não 70% ou 80%? Não foi possível encontrar uma resposta para isto na documentação disponível. Segunda nota técnica do Inep, os pesos

> “foram definidos após diversas reuniões técnicas com especialistas da área de educação superior. Posteriormente o assunto foi discutido na Comissão Nacional de Avaliação da Educação Superior (Conaes), que acabou por referendar os (seguintes) parâmetros”.

Foi, portanto, uma decisão discricionária, feita presumivelmente com a melhor das intenções, mas sem uma justificativa técnica ou educacional explícita. A decisão adotada significa dizer, aproximadamente, que o que os estudantes sabem ao final do curso (à medida que o Enade expressa isto) só caracteriza 40% da qualidade do curso. Isto contrasta com o antigo Exame Nacional do Curso (ENC, ou “Provão”), para o qual este componente (embora com diferenças importantes) era o único que importava, e tinha, portanto, um peso de 100%.[7] Uma das críticas feitas ao ENC é que ele não tomava em consideração o que os alunos aprendiam durante o curso, o conhecimento adicionado em relação ao que eles já tinham no início da formação. O IDD é uma tentativa de medir isto, mas, de novo, não há nenhuma razão explícita para que este componente tenha um valor de 30%, e não 20% ou 60%. Houve também a decisão de que os insumos deveriam valer 30%, embora a qualidade dos insumos já esteja refletida no segundo, como veremos mais adiante.

O Enade é resultante da combinação dos resultados dos diversos testes aplicados aos estudantes de primeiro e último ano dos cursos. O IDD é o “Indicador de Diferença entre os Desempenhos Observado e Esperado”, elaborado pelo Inep, por meio de análises estatísticas de regressão, com os dados do Enade. O componente de insumos é formado por quatro subcomponentes – infra-estrutura e instalações físicas, recursos didático-pedagógicos, docentes com doutorado e docentes em tempo integral, com pesos diferentes, que variam entre 10.2% a 38.9% dos 30% que eles representam. Apresentamos abaixo, de forma resumida, a maneira pela qual estes três componentes foram construídos, assim como alguns comentários a respeito.

---

[6] Ministério da Educação, Cálculo do Conceito Preliminar de Cursos de Graduação: Nota Técnica, Inep, 2007.

[7] Para uma análise detalhada do ENC, veja S. Scwhartzman, *The National Assessment of Courses in Brazil*. Disponível em http://www.schwartzman.org.br/simon/provao2.pdf

O Enade

A descrição detalhada dos procedimentos adotados para o Enade consta do documento "Resumo Técnico do Enade 2005", divulgado pelo Inep em 2006 como "versão preliminar sem revisão", e que aparentemente já não está disponível no "site" do Inep na Internet, e nem foi possível encontrar uma versão definitiva. É possível que alguns dos procedimentos adotados para o Enade 2005 tenham sido alterados posteriormente, mas isto não está documentado.
A nota do Enade é assim definida por este documento:

> A nota final da IES em um determinado curso é a média ponderada da nota padronizada dos concluintes no componente específico, da nota padronizada dos ingressantes no componente específico e da nota padronizada em formação geral (concluintes e ingressantes), considerando-se, respectivamente, os pesos 60%, 15% e 25%. Assim, a parte referente ao componente específico contribui com 75% da nota final, enquanto a referente à formação geral contribui com 25%, em consonância com o número de questões na prova, 30 e 10, respectivamente.

Existem três aspectos a considerar no Enade: as diferentes provas que são realizadas, os procedimentos de padronização, e a maneira pela qual estes diferentes dados são combinados em um conceito final.

O Enade consta de duas provas, uma de conhecimentos gerais e outra de conhecimentos específicos da área, que são aplicadas a uma amostra de alunos da primeira e da última série dos cursos. Para o cálculo do Enade, as médias das notas na prova específica dos alunos iniciantes e concluintes têm peso distinto (15% e 60%, respectivamente), enquanto que a média das notas da prova de formação geral é calculada em conjunto, tanto para iniciantes quanto para concluintes, e com o peso de 25%.

Existe uma série de questões relativas à maneira como o Enade é feito e calculado. As mais evidentes são as seguintes:

Medida de resultado final ou de valor adicionado? A realização de provas para alunos iniciantes e concluintes teria por objetivo medir o que o curso adiciona ao aluno em temos de conhecimento. Para avaliar isto, seria necessário subtrair o resultado da nota no primeiro ano do resultado da nota no último, supondo que a prova fosse a mesma. O Enade, no entanto, soma estas duas notas, dando um bônus aos cursos que já recebem alunos com melhor bagagem cultural e de informação, mesmo que os cursos adicionem pouco a isto. O IDD, elaborado posteriormente, foi uma tentativa do Inep de corrigir isto, mas esta ponderação continua sendo feita, aparentemente, porque se trata de uma antiga decisão da Comissão Nacional de Avaliação do Ensino Superior (Conaes) ainda não alterada.

O que mede a prova de conhecimentos gerais?

Esta prova, do ponto de vista técnico, contraria os princípios mais elementares de construção de testes e avaliações. Segundo o documento técnico do Enade,

> (...) no componente de avaliação da formação geral, dentro dos limites possíveis, é investigada a formação de um profissional ético, competente e comprometido com a sociedade em que vive. Nas questões da prova busca-se também obter indícios relativos à capacidade do estudante para analisar, sintetizar, criticar, deduzir, construir hipóteses, estabelecer relações, fazer comparações, detectar contradições, decidir e organizar as idéias. O componente de avaliação da formação geral do Enade 2005 foi composto por 10 (dez) questões, sendo 3 (três) questões discursivas e 7 (sete) de múltipla escolha, utilizando situações-problema, estudos de caso, simulações e interpretação de textos e imagens. As questões discursivas buscavam investigar, além do conteúdo específico, aspectos como a clareza, a coerência, a coesão, as estratégias argumentativas, a utilização de vocabulário adequado e a correção gramatical do texto. Na avaliação da formação geral buscou-se contemplar alguns dentre os vários temas propostos na legislação relativa ao exame, dentre eles: sociodiversidade, biodiversidade, globalização, novos mapas sociais, econômicos e geopolíticos, políticas públicas, redes sociais, relações interpessoais, inclusão e exclusão digital, cidadania, além de outros problemas contemporâneos.

É simplesmente impossível medir todas estas coisas com três questões discursivas e sete questões de múltipla escolha. Para ter um mínimo de validade, uma prova que quisesse medir tudo isto deveria ter vários itens para cada uma das 25 competências listadas acima, cada um deles devidamente testado e validado. Feita sem qualquer metodologia clara, a prova se reduz, na prática, a uma medida impressionista de cultura geral, tal como interpretam as pessoas responsáveis pela sua avaliação.

O que medem as provas de conhecimento específico?

As provas de conhecimento específico são elaboradas por professores das respectivas disciplinas, e os alunos recebem notas de 0 a 100, que depois são padronizadas e reduzidas a uma escala de 0 a 5, e finalmente agrupadas em uma escala de 5 pontos. O Inep divulga as médias das turmas por curso em uma escala de 0 a 100, e podemos supor que elas refletem o que os professores entendem por bons ou maus resultados em uma escala convencional. Se as escalas fossem comparáveis entre os cursos, a fonoaudiologia seria a melhor área na avaliação de 2007, com média de 63.1 pontos, seguida da de medicina. Mas não temos como saber se os professores que prepararam as provas de medicina têm o mesmo nível de exigência dos de fonoaudiologia ou zootecnia, que também tem média acima de 60, e o MEC em nenhum momento explicita quais seriam os níveis de conhecimentos e competências considerados satisfatórios para cada uma das áreas.

Ao que tudo indica, as provas de conhecimento específico, assim como o componente geral, não atendem a nenhum dos requisitos técnicos esperados para avaliações deste tipo.

O Ministério da Educação, em várias portarias do dia 07 de agosto de 2008 (disponíveis no "site" do Inep), procurou definir os conteúdos para o Enade de 2008. Para o componente geral, as portarias dizem que "as questões versarão sobre *alguns (sic)* dos seguintes temas", ao que se segue uma lista de 20 itens, dentre os quais "exclusão e minorias", "arte, cultura e filosofia", "globalização" e "propriedade intelectual", além de oito capacidades (como "ler e interpretar textos" e "questionar a realidade") e cinco competências (como "construir perspectivas integradoras"). Tudo isto com sete questões de múltipla escolha e três questões discursivas.

As provas de conhecimento específico, em geral, consistem em listas de competências e matérias curriculares elaboradas por comissões de especialistas convidados das diferentes áreas. Como ilustração, a prova de filosofia consiste em seis competências (como a "capacidade de desenvolver uma consciência crítica sobre conhecimento, razão e realidade sócio-histórico-política" e "percepção da integração necessária entre a filosofia e a produção científica, artística, bem como com o agir pessoal e político", e 30 questões de conteúdo, dentre as quais "tempo e eternidade, conhecimento humano e conhecimento divino" e "crítica à metafísica na contemporaneidade. Nietzsche. Wittgenstein. Heidegger". Para a prova de química, são listadas dezenove habilidades e competências gerais, dentre as quais "conduzir processos investigativos em todas as suas etapas compreendendo a elaboração de projetos, sua execução, comunicação e socialização de resultados" e uma grande lista de conhecimentos específicos. Não há nenhuma indicação sobre o peso relativo que as diferentes competências e conhecimentos devem ter na avaliação, e menos ainda sobre como estas questões serão transformadas em itens de prova, que possam resultar em uma prova nacional de 30 questões. Tudo isto, naturalmente, supondo que existe clareza sobre as competências específicas das diferentes áreas, assim como sobre os conhecimentos que devem ter prioridade.

À medida que estas provas tenham alguma validade, a comparação entre as médias dos alunos iniciantes e dos concluintes nos componentes específicos poderia ser uma medida da qualidade de um curso – o que eles efetivamente acrescentam ao que os alunos já trazem em sua bagagem. Esta

Área	Iniciantes	Concluintes	Diferença no curso
Agronomia	37.86	54.85	17.47
Biomedicina	28.74	42.32	13.25
Educação Física	41.71	51.74	9.87
Enfermagem	25.57	37.38	11.72
Farmácia	31.03	42.03	10.58
Fisioterapia	32.64	50.77	18.28
Fonoaudiologia	44.12	63.12	19.04
Medicina	32.49	61.21	29.38
Medicina Veterinária	27.13	50.16	22.56
Nutrição	35.13	49.35	14.37
Odontologia	31.78	55.06	23.22
Serviço Social	35.16	43.64	10.43
Tecnologia de Radiologia	19.39	29.94	9.71
Tecnologia em Agroindústria	27.27	45.73	15.12
Terapia Ocupacional	27.59	35.63	0.87
Zootecnia	37.12	61.61	25.64
TOTAL	32.88	48.38	15.25

Figura 1 – Médias e diferenças das provas de conhecimento específico, por área (fonte: Enade 2007, dados do Inep).

comparação dos dados de 2007 mostra que os cursos de medicina, seguidos dos de zootecnia e odontologia, são os que aparentemente mais acrescentam competências aos alunos, enquanto que os de radiologia, terapia ocupacional e educação física seriam os que menos acrescentam. É esta dimensão de valor adicionado que o IDD vai procurar captar.

Como são feitas a padronização e as ponderações dos resultados? Para poder combinar os resultados das diferentes provas, o Inep recorre a procedimentos estatísticos de padronização dos resultados. Estes procedimentos incluem calcular a média dos resultados de cada turma; colocar a média dos cursos em termos de seu afastamento, em desvios padrão, das médias dos cursos dos setores; transformar esta escala padronizada, que deve variar entre –3 e +3, com média zero, em uma escala de 0 a 5; e finalmente agrupar os resultados obtidos em cinco categorias.

É difícil dizer que conseqüências estas transformações têm para a interpretação final dos resultados, mas cabem pelo menos duas observações. Como a média dos alunos iniciantes no componente específico é padronizada, em relação ao conjunto de iniciantes, e a média dos concluintes, em relação ao conjunto dos concluintes, torna-se impossível examinar as diferenças entre o desempenho dos iniciantes e concluintes depois da padronização. Depois, a transformação para a escala de 0 a 5 deveria resultar na média para todas as áreas de 2.5 em cada prova, mas, como a transformação é feita por uma fórmula baseada nos valores mínimos e máximos de cada área, que variam, as médias também variam. Finalmente, o Inep não apresenta os resultados em uma escala contínua de 0 a 5, e sim em uma escala discreta de 5 pontos, pelo agrupamento dos resultados, que deveria ter uma média igual a três.

São estes valores padronizados das médias das três provas – componente geral para todo o curso, e componente específico para ingressantes e concluintes – que são combinados, com pesos diferentes, para definir o valor do Enade. Como é que o peso destas provas é definido? O documento técnico mencionado anteriormente diz que os pesos são atribuídos “em consonância com o número de questões na prova, 30 e 10, respectivamente”, o que não faz sentido, já que o número de questões em uma prova é uma questão meramente operacional, e não diz nada a respeito de sua importância ou peso relativo. Ao atribuir peso de 25% para a prova de competência geral o Enade aumenta ainda mais o bônus de pontos dados às instituições capazes de atrair alunos com mais cultura geral e capacidade de escrever, normalmente de condições socioeconômicas mais altas, ficando somente 60% do conceito para os resultados obtidos pelos alunos ao final do curso (os outros 15%, inexplicavelmente, vêm dos conhecimentos específicos que os alunos iniciantes trazem para o curso). Isto significa que, hipoteticamente, um curso que tem nota máxima, 100, no exame final, mas recebe alunos de origem social mais baixa, sem nenhum conhecimento prévio de sua área de especialização e pouca cultura geral, ficaria com 60 no Enade, em uma escala de 0 a 100; igual a um curso que só tivesse nota 20 na prova final, mas recebesse alunos de classe social mais alta, com mais cultura geral, que começariam com 40% da nota.

Ao final, ao invés de as notas do Enade terem uma distribuição normal, com a mesma média por área de conhecimento, estas médias variam, não em função da qualidade dos cursos, mas das

transformações realizadas com os dados iniciais, como se pode ver no quadro que se segue. A conseqüência é que uma área como fonoaudiologia acaba tendo uma média alta, de 3.8, que coloca boa parte de seus cursos em nível aparentemente alto de desempenho, enquanto que a de enfermagem, com 2.67, na média, fica prejudicada. Isto torna ainda mais difícil interpretar o que os conceitos finais do Enade realmente significam.

Conceitos do Enade após ajustes

Área	Média	N	Desvio Padrão	Mínimo	Máximo
Agronomia	3.40	104	1.02	1.00	5.00
Biomedicina	2.18	60	1.13	1.00	5.00
Educação Física	3.13	343	0.78	1.00	5.00
Enfermagem	2.67	338	0,87	1.00	5.00
Farmácia	2.46	190	1.05	1.00	5.00
Fisioterapia	3.10	274	0.74	1.00	5.00
Fonoaudiologia	3.82	56	0.94	1.00	5.00
Medicina	3.00	103	1.09	1.00	5.00
Medicina Veterinária	2.81	104	1.02	1.00	5.00
Nutrição	2.42	159	0.92	1.00	5.00
Odontologia	3.18	149	1.01	1.00	5.00
Serviço Social	3.56	133	0.96	1.00	5.00
Tecnologia de Radiologia	2.58	31	1.09	1.00	5.00
Tecnologia em Agroindústria	2.57	7	1.13	1.00	4.00
Terapia Ocupacional	3.59	29	0.95	1.00	5.00
Zootecnia	3.35	48	1.06	1.00	5.00
TOTAL	2.96	2.128	1.00	1.00	5.00

Figure 2 – Calculado a partir dos dados divulgados pelo Inep.

A representatividade do Enade

Enquanto que, no antigo ENC, todos os alunos na última série dos cursos avaliados tinham a obrigação de participar, o Enade é aplicado a uma amostra de alunos, o que tem levantado dúvidas sobre se esta amostra é efetivamente representativa das turmas. Para evitar problemas de seleção, a atual legislação torna obrigatória a inscrição dos alunos dos cursos sendo avaliados no Exame, o que permite que os que façam a prova sejam posteriormente sorteados de forma aleatória. Este procedimento, em princípio, deve evitar problemas de seleção tendenciosa, embora não existam informações disponíveis a respeito. Outro problema, comum tanto ao antigo ENC quanto ao Enem, é que o aluno pode comparecer ao exame e deixar a prova em branco, pois o que fica registrado no seu currículo é o comparecimento.

Conclusões sobre o Enade

O Enade é um exame com importantes limitações, das quais se destacam:

- A prova de conhecimentos gerais não tem nenhuma qualidade técnica, e seus resultados só podem ser interpretados, eventualmente, como refletindo a cultura geral dos estudantes e sua capacidade de escrever, nunca as inúmeras dimensões que ela deveria avaliar.

- As provas de conhecimentos específicos também não foram submetidas, aparentemente, a procedimentos de validação, exceto pela eliminação de itens de baixa correlação com os demais. A lista de competências e conhecimentos a serem avaliados é muito maior do que o que pode ser avaliado em um exame de 30 itens, e não existem critérios para estabelecer que resultados são considerados satisfatórios, a não ser a simples posição dos cursos nas distribuições.

- A aplicação das provas aos alunos iniciantes e concluintes, que deveria medir o valor adicionado dos cursos aos conhecimentos que os estudantes trazem, acaba sendo usada no sentido oposto, já que seus resultados são somados aos das provas dos concluintes. Isto, e mais a soma dos resultados dos alunos nas provas de conhecimentos gerais, faz com que os resultados do Enade sejam tendenciosos, no sentido de aumentar as pontuações dos cursos que recebem alunos com níveis mais altos de conhecimento e informação – geralmente de nível socioeconômico mais alto.

- Ao utilizar uma prova única para cada área de conhecimento, o Enade, da mesma forma que o antigo ENC, impõe um padrão único de ensino a todas as instituições, prejudicando, assim, a diversidade que deveria ser respeitada e valorizada.
- Os procedimentos estatísticos de estandardização das notas são confusos, e os critérios de ponderação das diferentes provas são discricionários, o que torna a interpretação dos resultados do Enade ainda mais incerta.

O Índice de Diferença de Desempenho (IDD)

Este índice foi elaborado pelo corpo técnico do Inep com os dados do Enade de 2005, em um esforço para recuperar a intenção inicial do exame de avaliar a capacidade dos cursos em adicionar competências a seus estudantes, a partir da bagagem de informações e competências com a qual iniciam seus estudos, intenção que se viu frustrada pela decisão, aparentemente do Conaes, de somar os resultados das avaliações iniciais e finais. No documento técnico da *Pesquisa Nacional por Amostra de Domicílios* (PNAD) de 2005 o IDD é definido como uma medida da diferença entre

O desempenho médio obtido no Enade pelos concluintes de cada curso é o desempenho médio que seria esperado ao final do curso para o perfil de ingressantes daquela instituição caso eles tivessem freqüentado um curso de qualidade correspondente à média dos cursos que participaram do Enade na mesma área e que possuam ingressantes com perfil similar.

O cálculo do IDD

Os detalhes da estimativa dos valores do IDD são complexos. Ela é feita por análise de regressão, e está disponível na nota técnica do Inep que estamos utilizando. A idéia geral é que, se os estudantes iniciam o ensino superior em determinadas carreiras com determinadas características, em função de seu nível socioeconômico ou da educação de seus pais, eles deveriam chegar a determinados níveis de desempenho ao final do curso, que é estimado a partir do desempenho médio dos ingressantes com suas características na área. Estudantes com menos condições chegariam a resultados menores do que os de melhores condicionais iniciais, mas o seu ganho relativo poderia ser maior. O que o IDD faz é comparar o desempenho dos estudantes que se formam (definido por uma combinação da prova de conhecimentos específicos com a prova de conhecimentos gerais, com os pesos de 75% e 25%, respectivamente) com o desempenho médio esperado dos estudantes com características iniciais similares.

Valores Médios de IDD por área de conhecimento e setor

Área	Privada	Pública	Total
Agronomia	2.56	2.77	2.69
Biomedicina	2.78	3.00	2.81
Educação Física	2.96	3.12	3.00
Enfermagem	2.90	3.58	3.04
Farmácia	3.22	3.50	3.28
Fisioterapia	2.72	3.36	2.78
Fonoaudiologia	2.88	3.33	2.90
Medicina	2.79	3.17	2.97
Medicina Veterinária	2.94	3.35	3.08
Nutrição	2.85	3.64	2.97
Odontologia	2.82	3.52	3.05
Serviço Social	2.95	3.48	3.06
Tecnologia de Radiologia	2.80	3.00	2.83
Terapia Ocupacional	3.47	2.33	3.32
Zootecnia	2.10	2.75	2.56
TOTAL	2.89	3.27	2.98

Figure 3 – Calculado a partir dos dados divulgados pelo Inep.

No relatório do Enade de 2005, os dados do IDD foram apresentados de forma padronizada, com média zero e desvio padrão de um por área de conhecimento.[8] Para o cálculo do Conceito Preliminar, os dados foram colocados em uma escala de 0 a 5 e depois em cinco categorias, aparentemente com os mesmos problemas de ajustes que ocorreram com a padronização do Enade. O resultado é que, ao invés de todas as áreas terem média três, os valores apresentam diferenças por áreas de conhecimento, dando a impressão de que algumas são melhores do que outras em sua capacidade de adicionar conhecimentos. A comparação entre instituições públicas ou privadas dentro de cada área de conhecimento, no entanto, pode ser feita, e os dados mostram que, de uma maneira geral, as instituições públicas acrescentam mais conhecimento do que as privadas, com diferenças maiores ou menores nas diferentes áreas.

Conclusões sobre o IDD

As principais conclusões a respeito do IDD são as seguintes:

- A elaboração do IDD é um exercício analítico interessante, que avança sobre os resultados convencionais do Enade;

- Não é clara a razão pela qual o desempenho final é calculado por uma combinação de 75% da prova de conhecimento específico e 25% da prova de conhecimento geral, já que esta última não tem interpretação plausível, e simplesmente reflete as condições iniciais de cultura geral dos alunos. Isto pode estar reduzindo, aparentemente, o efeito observável da capacidade dos cursos de adicionar conhecimentos aos alunos;

- O IDD é uma estimativa estatística, e como tal não pode ser transposto mecanicamente para a avaliação de cursos específicos.

O cálculo dos insumos e sua ponderação

A maneira pela qual este componente é calculado, e a maneira pela qual os pesos utilizados foram gerados, estão descritos no documento *Cálculo do Conceito Preliminar de Cursos de Graduação – Nota Técnica* (Inep, 2008).

---

8 Sistema Nacional de Avaliação da Educação Superior (Sinaes), Resumo Técnico Enade 2005, Brasília, Ministério da Educação, 2006.

Essencialmente, o que fez o Inep foi estimar em que medida os dados sobre qualidade da infra-estrutura e recursos didáticos pedagógicos dos cursos, tal como percebidos pelos alunos e o número de professores com doutorado e em tempo integral se relacionam com o IDD, e usar isto como critério para as ponderações.

A qualidade da infra-estrutura e os recursos didáticos pedagógicos foram estimados a partir das respostas dos estudantes ao questionário que preencheram para a prova do Enade. Não está claro como os dados sobre número de professores com doutorado e em tempo integral foram obtidos. O Censo do Ensino Superior não tem esta informação por curso, e sim por instituição. Isto ocorre porque, em sua maioria, as universidades estão organizadas por departamentos e não cursos, e os professores de determinado curso podem vir de departamentos distintos. Existe um cadastro de professores com informações por curso fornecidas pelas instituições que é, aparentemente, a fonte dos dados utilizados. Mas como, na realidade, esta distribuição por curso nem sempre existe, a qualidade desta informação é questionável.

Para estimar o peso de cada um destes componentes no índice de insumo, foi feita uma análise de regressão, tendo como variável dependente o IDD, que o Inep considera "como *proxy* de qualidade". A regressão busca medir, estatisticamente, em que medida a qualidade da infra-estrutura, os recursos didáticos, a quantidade de professores em tempo integral e o número de doutores afetam o IDD. Não há nenhuma explicação no documento sobre a razão desta escolha do IDD, e não do Enade, como variável dependente, ainda que se possa imaginar que tenha sido por causa da maneira em que os diferentes componentes do Enade são combinados, como descrito acima. A escolha das variáveis independentes foi feita a partir do nível de significação que apresentaram em análises de regressão.

A análise de regressão permite produzir uma equação em que a variável dependente, o Índice de Diferença de Desempenho, é explicada pela combinação das variáveis independentes, com os respectivos pesos, mais um fator aleatório, que dá conta de tudo o mais que não está incluído nas variáveis independentes.

$$IDD_i = \beta_1 Infra_i^{padr} + \beta_2 Pedag_i^{padr} + \beta_3 Dout_i^{padr} + \beta_4 Integ_Parc_i^{padr} + \varepsilon_i$$

Segundo o documento do Inep,

> O modelo acima foi estimado a partir dos resultados do Enade e IDD dos anos de 2004, 2005 e 2006 (primeiro ciclo de avaliação) e do Cadastro de Docentes do Ensino Superior do ano-base de 2006. A idéia é de que os pesos descritos na Tabela 1 são fixos e serão aplicados aos resultados de outros anos de avaliação para a construção do Conceito Preliminar de Cursos de Graduação nos anos subseqüentes.

Tabela 1 – Resultados do modelo e pesos de cada Insumo no IDD

Variável Dependente = 'IDD'	Coeficiente (desvio padrão)	Peso do Atributo	
Infra^{padr} = 'aulas práticas, os equipamentos disponíveis são suficientes para o número de estudantes'	0,387* (0,0139)	10,2%	
Pedag^{padr} = 'os planos, de ensino contêm todos os seguintes aspectos: objetivos; procedimentos de ensino e avaliação; conteúdos e bibliografia da disciplina'	0,1023* (0,0137)	27,1%	
Dout^{padr} = 'percentual de professores (no mínimo) doutores no curso'	0,1472* (0,0129)	39,9%	62,79%
Integ_parc^{padr} = 'percentual de professores que cumprem regime parcial ou integral (não-horista) no curso'	0,0898* (0,0133)	23,8%	

Nº Obs: 7666 cursos

F (4,7662) = 80,9

Prob > F = 0.000

Nota: Coeficiente estatisticamente significativo a 1%

Os resultados estão no quadro acima. Os pesos da última coluna foram calculados por simples regra de três a partir dos coeficientes da segunda coluna. Assim, o peso do percentual de doutores significa que seu coeficiente, 0.1472 (que é o $\beta 3$ na equação acima) representa 38,9% da soma dos quatro β. A nota técnica não indica o valor do componente aleatório, não explicado por estes quatro componentes, nem o coeficiente de correlação múltiplo resultante da regressão, e por isto ficamos sem saber quanto que o IDD é explicado por estes fatores ou por outros que não aparecem ou não foram observados. A única informação disponível é que a relação destes quatro componentes com o IDD é estatisticamente significativa.

Além destes procedimentos, a nota técnica descreve uma série de transformações efetuadas para padronizar os conceitos, para colocá-los em escalas comparáveis de cinco pontos, e depois combiná-los, semelhantes aos utilizados para a preparação dos índices do IDD e do Enade.

Conclusões sobre o Índice de Insumos

- Índice de Insumos é baseado em uma estimativa estatística de precisão desconhecida, e por isto o uso de pesos com uma casa decimal dá uma impressão falsa de certeza.

- Os dados sobre professores de tempo integral e em dedicação exclusiva, salvo melhor informação, que são geralmente da instituição, e não dos cursos, são precários.

- Não há explicação sobre a escolha do IDD, e não do Enade, como referência para a construção do Índice de Insumo.

Os resultados do Conceito Preliminar são tendenciosos?

A incorporação das informações sobre professores com doutorado e em tempo integral fez com que surgisse a dúvida se o Conceito Preliminar não estaria introduzindo uma tendência favorável às instituições públicas. A comparação dos dados do CPC com seus componentes, assim como com a prova de conhecimentos específicos dos formandos, não mostra esta tendenciosidade. Na média, o IDD é marginalmente menor que o Enade das instituições públicas, e um pouco maior nas instituições privadas. Os dados do índice de insumo por instituição não foram disponibilizados pelo Inep.

	Instituições privadas				Instituições públicas			
Área	CPC	Conhecimentos específicos	IDD	Enade	CPC	Conhecimentos	IDD	Enade
Agronomia	2.9	3.4	2.6	3.1	3.4	3.9	2.8	3.6
Biomedicina	2.5	3.0	2.8	1.7	3.7	4.0	3.0	3.8
Educação Física	2.9	3.3	3.0	3.1	3.2	3.5	3.1	3.3
Enfermagem	2.7	2.6	2.9	2.4	3.6	3.6	3.6	3.6
Farmácia	2.7	2.1	3.2	2.1	3.5	3.6	3.5	3.6
Fisioterapia	2.9	3.1	2.7	3.0	3.4	3.8	3.4	3.8
Fonoaudiologia	3.4	4.3	2.9	3.9	3.9	3.8	3.3	3.7
Medicina	2.7	3.9	2.8	2.7	3.3	4.2	3.2	3.3
Medicina Veterinária	2.7	2.1	2.9	2.2	3.5	3.5	3.4	3.8
Nutrição	2.7	2.9	2.9	2.2	3.6	3.8	3.6	3.4
Odontologia	2.8	3.1	2.8	2.7	3.8	4.5	3.5	4.2
Serviço Social	3.2	3.8	3.0	3.7	3.3	3.3	3.5	3.3
Tecnologia de Radiologia	2.6	2.3	2.8	2.2	4.0	4.0	3.0	4.4
Tecnologia em Agroindústria		2.0			3.0	3.4		2.6
Terapia Ocupacional	3.3	3.4	3.5	3.7	3.8	3.0	2.3	3.2
Zootecnia	2.3	3.0	2.1	2.6	3.4	3.7	2.8	3.7
TOTAL	2,8	3,0	2.9	2.7	3.5	3.7	3.3	3.6

O Conceito Preliminar e as melhores práticas

O Conceito Preliminar é uma construção estatística baseada em uma série de aproximações e pressupostos não explicitados, que podem fazer sentido como exercício de análise e até mesmo para ajudar nas decisões internas por parte do Ministério da Educação, mas nunca poderiam ser difundidos publicamente como avaliações da qualidade dos cursos, ainda que com o título de "preliminar". A análise detalhada do índice mostra que, além das questões mais técnicas já indicadas, existem as seguintes questões de natureza mais substantiva:

- O Enade aumenta artificialmente a pontuação dos cursos que recebem alunos de maior nível educacional;

- IDD foi construído para compensar este problema com o Enade;

- O Índice de Insumo, por ser estimado em função do IDD, significa na prática uma duplicação do IDD, à medida que esta estimativa seja completa. Como não o é (e a proporção da variância do IDD explicada pelos insumos não está dita), isto significa simplesmente atribuir mais qualidade a cursos localizados em instituições geralmente públicas que podem dispor de mais professores em tempo integral e com doutorado.

- As opiniões dos alunos sobre os recursos e a qualidade didático-pedagógica dos cursos são uma informação precária, já que estas avaliações dependem sempre de referências e comparações subjetivas que podem variar muito de lugar a lugar e pessoa a pessoa. Estas opiniões podem ser úteis como coadjuvantes para uma análise global de um curso ou instituição, mas não podem ser consideradas um indicador sólido, e muito menos com um peso tão preciso como 27,1% no índice que o Inep adota.

- Ao combinar duas medidas em boa parte contraditórias (o Enade, com peso de 40%, e o IDD, com peso de 60% - já que o índice de insumo é calculado a partir de sua equivalência com o IDD), o conceito preliminar torna-se ainda mais confuso e de difícil interpretação.

As práticas de avaliação externa são hoje quase universais, e existe uma instituição internacional, a Rede Internacional de Agências de Asseguramento de Qualidade da Educação Superior (*International Network for Quality Assurance Agencies in Higher Education*, Inqaahe), que reúne mais de 200 agências de avaliação do ensino superior em todo o mundo, e que tem por objetivo compartir as experiências e incentivar as melhores práticas a respeito. No "*policy statement*" do Inqaahe sobre avaliações externas é possível destacar os seguintes pontos principais:

- *Primazia da liberdade acadêmica e da integridade:* A educação superior pode se dar de muitas formas, mas a Rede considera que as características que definem a educação superior incluem políticas claras e compromisso com a integridade e liberdade acadêmica, que se baseiam no reconhecimento de que as atividades acadêmicas devem ser conduzidas em um espírito de honestidade e abertura. Avaliações externas devem ser conduzidas de modo tal que promovam a liberdade acadêmica e a integridade institucional.

- *As instituições de nível superior são as principais responsáveis pelas avaliações:* As atividades de avaliação externa devem ser baseadas na premissa que a qualidade e a avaliação da qualidade são da responsabilidade, em primeiro lugar, das próprias instituições, e que os processos de avaliação externos devem respeitar a integridade institucional.

- *Independência das avaliações:* Todos os esforços devem ser feitos para garantir a independência dos avaliadores individuais, incluindo a adoção de políticas claras sobre conflitos de interesse reais ou potenciais. As avaliações externas devem ser desenvolvidas em um contexto em que seus critérios e procedimentos sejam publicados com antecipação às avaliações. A formulação destes critérios e procedimentos devem incluir consulta com os principais interessados. Os processos de avaliação externa devem ser desenvolvidos em um espírito de independência, e os julgamentos resultantes das avaliações não devem sofrer quaisquer tipos de influência, de governos, instituições, ou de outras partes.

- Em algumas circunstâncias, as agências de avaliação não têm poderes de decisão, mas fazem recomendações para outras. Nestes casos, é importante que as agências possam fazer recomendações independentes que, exceto em casos excepcionais, devem ser públicas.[9]

À luz das melhores práticas, o "conceito preliminar" preparado e divulgado pelo Ministério da Educação deixa muito a desejar:

---

[9] *Primacy of Academic Freedom and Integrity: While higher education comes in many forms, the Network holds that the defining characteristics of higher education include clear policy and procedural commitments to academic integrity and academic freedom, which is the recognition that academic endeavors should be wholly conducted in a spirit of honesty and openness. External quality assurance should be conducted in such a way as to promote academic freedom and intellectual and institutional integrity. Higher Education Institutions (HEIs) have prime responsibility for Quality Assurance: EQA activities should be based on the premise that Quality and Quality Assurance are primarily the responsibilities of HEIs themselves and should respect institutional integrity.Independence of Evaluation: Every effort should be made to ensure the independence of the individual evaluators, including the adoption of a clear policy on real or potential conflicts of interest. EQA (External Quality Assessment) should be carried out in the context of criteria and procedures that are published in advance of any review. The formulation of such criteria and procedures should include consultation with major stakeholders. The EQA process itself should be carried out in a spirit of independence, and judgments made as part of the review should not be subject to external influences whether from governments, institutions or elsewhere. It is accepted that in some circumstances an EQA agency does not itself have decision- making powers but makes recommendations to others. In such cases it is important to ensure that the agency is in a position to make an independent recommendation which, save in exceptional circumstances, is made public. (http://www.inqaahe.org/).*

- Não tem legitimidade, porque não foi elaborado com a participação e o envolvimento de setores relevantes da comunidade de ensino superior do país, que foram surpreendidos com sua divulgação.

- Não é uma avaliação independente, porque foi realizada por uma agência governamental, que tem suas próprias orientações e preferências.

- Não é inteligível, porque a forma pela qual o "conceito preliminar" foi produzido não é clara para a maioria das pessoas, e os documentos técnicos disponíveis não dão informações suficientes, e são incompreensíveis para as pessoas sem formação estatística.

- Comete o erro de transformar indicadores "*low stakes*", como o IDD e o índice de insumo, produzidos por processos de estimações estatísticas, em um conceito de altas conseqüências, que, ao ser publicado, afeta o prestígio e a reputação das instituições eventualmente mal-avaliadas, enquanto que atribui qualidades a instituições que talvez não a tenham;

- Não toma em consideração os preceitos centrais do "*policy statement*" do Inqaahe, à medida que: os procedimentos adotados não foram produzidos em consulta com as instituições avaliadas; a avaliação não foi feita de forma independente; e a divulgação dos resultados foi feita de forma intempestiva, que pode ter afetado de forma bastante grave a reputação de muitas instituições.

Idéias para desenvolver um sistema mais adequadode avaliação do ensino superior no Brasil

A questão de como regular e avaliar a educação superior no Brasil precisa ser discutida com mais profundidade. Como sugestão, é possível listar algumas alternativas possíveis.

- *Criar uma agência de regulação do ensino superior autônoma.* A experiência brasileira com as agências de regulação nem sempre é bem-sucedida, porque corre o duplo risco de ser capturada pelos grupos e interesses do setor regulado, e sofrer interferência política e partidária na escolha de seus dirigentes. No entanto, uma agência autônoma bem-concebida e formada por pessoas de alta reputação e que não sejam, simplesmente, porta-vozes das instituições reguladas ou do governo, poderia dar mais legitimidade a todo este processo. Uma agência como esta deveria substituir o atual Conaes.

- *Descentralizar as avaliações.* Nenhuma agência nacional tem condições de desenvolver um sistema de avaliação que possa cobrir, de forma adequada e não burocrática,

todas as 2.270 instituições e mais de 22 mil cursos superiores presenciais que existem no Brasil, sem falar nas novas modalidades de cursos a distância, semipresenciais etc. Não adianta descentralizar para os estados, ou regionalmente, porque as estruturas governamentais na maioria dos estados são ainda mais precárias que as do Ministério da Educação. A melhor solução é permitir que se criem agências de avaliação de direito privado que sejam credenciadas pela agência reguladora nacional. A legislação deveria requerer que todas as instituições estivessem filiadas a uma destas agências e contribuíssem financeiramente para sua manutenção.

- *Criar um mercado competitivo de qualidade.* Além de agências de credenciamento institucional, poderia haver agências e credenciamento por área de conhecimento, com a participação das organizações profissionais e acadêmicas correspondentes. Para as profissões que afetem a saúde ou o patrimônio das pessoas, deve ser estimulada e fortalecida a certificação individual, na linha do Exame de Ordem da Ordem dos Advogados do Brasil (OAB). As agências de credenciamento e certificação não seriam monopólios, mas poderiam competir entre si.

- *Substituir as notas ou conceitos por certificações.* O governo federal deveria desistir do propósito de distribuir "estrelas" de qualidade para as instituições de ensino superior e se dedicar simplesmente, com o apoio das agências de certificação, a declarar se as instituições e cursos estão aptos ou não a exercer as atividades a que se propõem. O sistema de "estrelas" tem se mostrado muito confuso, difícil de entender, e acaba impondo padrões únicos e rígidos às instituições, que tolhem sua liberdade e iniciativa e tornam ilegítima a diferenciação.

- *de conhecimento*. Um dos efeitos positivos do antigo Exame Nacional de Cursos era que os professores das diferentes áreas de conhecimento eram mobilizados para elaborar a prova e, depois, para discutir e avaliar os resultados. É possível que o mesmo esteja ocorrendo com o Enade. Para que estes esforços sejam frutíferos, é necessário que eles sejam feitos de forma tecnicamente competente, e que resulte na identificação do leque de competências centrais das diversas áreas de formação, permitindo assim que os diversos cursos superiores escolham e tornem claras as competências que se disponham a proporcionar.

O papel do Ministério da Educação e do setor privado

O Ministério da Educação tem a responsabilidade de zelar pela qualidade da educação superior como um todo, mas não lhe cabe o papel nem o direito de definir o "que" e "como" as instituições devem ensinar. A educação não é, como às vezes se interpreta, uma função ou conces-

são do Estado, mas um direito das pessoas, que cabe ao setor público assegurar e apoiar. Ao mesmo tempo, o Ministério da Educação mantém uma rede própria bastante significativa de instituições financiadas com recursos públicos, e é de sua responsabilidade zelar para que elas funcionem da maneira mais competente e eficiente possível. Além disto, o Ministério mantém programas de subsídios diretos e indiretos ao setor privado, dos quais o mais destacado é o Programa Universidade para Todos (ProUni), e deveria ter a responsabilidade de garantir que estes programas subsidiados funcionem a contento. Seria importante que o Ministério da Educação, ao invés de dar prioridade à avaliação das instituições privadas que não financia, concentrasse seus esforços em avaliar suas próprias atividades, também com o auxílio de avaliadores externos.

A questão do quanto o setor privado precisa ser avaliado e regulado pelo setor público é polêmica, e precisa ser aprofundada. Existindo bons sistemas de certificação profissional para áreas como medicina, odontologia e direito, a avaliação dos cursos superiores torna-se menos premente. Nas outras áreas, e mesmo na área do direito, em que só uma minoria dos formados consegue ser aprovada no Exame da Ordem dos Advogados do Brasil (OAB), é possível argumentar que, de alguma maneira, as pessoas estão se beneficiando da educação que recebem, e que isto pode ser deixado à regulação do próprio mercado. É possível argumentar também, por outro lado, que, dada a precariedade das informações existentes sobre a qualidade dos cursos, e o tempo que transcorre entre o início de um curso e a tentativa do estudante de entrar no mercado de trabalho, é importante que as pessoas tenham informações para fazer escolhas adequadas, o que justifica, pelo menos, um sistema de certificação inicial.

O setor privado, por outra parte, tem se beneficiado muito da presunção de que os diplomas que emite são equivalentes aos das melhores instituições do país. Esta presunção é um dos fatores que explicam a proliferação de "fábrica de diplomas" com pouco ou nenhum conteúdo, mas que as pessoas procuram, como procuram os cartórios, como forma de melhorar sua qualificação formal no mercado de trabalho, sobretudo no setor público. Os sistemas de avaliação existentes até aqui trataram de lidar com este problema, mas sem muitos resultados. À medida que se fortaleçam os sistemas de certificação profissional e institucional, e que se difundam as informações sobe a qualidade dos cursos em função se seus objetivos, a competição por qualidade tenderá a se fortalecer e a educação burocrática e cartorial perderá espaço, para o benefício de todos.